人工知能学会研究会資料
SIG-SLUD-B402-05

ソーシャルメディアにおける対話エージェントとユーザのコミュニケーション分析

An Analysis of Human-Agent communication in Twitter

稲葉 通将 [1*]　高橋 健一 [1]
Michimasa INABA[1]　Kenichi TAKAHASHI[1]

[1] 広島市立大学大学院情報科学研究科
[1] Graduate School of Information Sciences, Hiroshima City University

Abstract: In this paper, we create a non-task-oriented dialogue agent "KELDIC" on Twitter, and anlyze communication between the agent and twitter users. On twitter, users can react to tweets from others in three ways, reply, adding to favorite and retweet. From the point of view of user reactions to KELDIC's tweets, we demonstrate statistical features of users' behaviour. The result of analysis indicates the possibility of quantitative evaluation of dialogue agents using users' reaction.

1　はじめに

人間とオープンドメインな対話を行うことができる非タスク指向型対話エージェントは，エンターテイメント用途のみならず，認知症の緩和やカウンセリングなど様々な場面での活用が期待されており，注目が集まっている [1].

最近では，Twitter などソーシャルメディアのデータを用いた非タスク指向型対話エージェントの研究が活発に行われている．例えば，Twitter におけるツイート・リプライのペアを大量に収集しておき，ユーザの入力に類似したツイートを検索し，それに対するリプライをシステムの応答として使用する手法 [2] や，Twitter 上の対話から対話モデルを構築した研究 [3]，Twitter 上の対話と映画の脚本を対話データとして用いてエージェントを構築した研究 [4] などがある．ソーシャルメディアのデータは，非タスク指向型対話エージェントが対象としているオープンドメインな対話を多分に含んでいることに加え，WebAPI の整備などにより容易に大量のデータが取得可能であり，当該分野との親和性は高い．前述した既存の研究では，ソーシャルメディアは対話データを取得するための場として利用していた．しかし，対話エージェントの対話相手となるユーザを集めるためのコストが不要であり，また幅広い年代の様々な趣味嗜好を持つユーザが存在することから，エージェントが実際に対話を行う場としてもソーシャルメディアは適していると思われる．さらに，ソーシャルメディア上の対話エージェントに対するユーザの振る舞いは，エージェントの応答内容に大きく左右されると考えられることから，その情報を用いることで，エージェントの性能自動評価や対話の破綻検出が行える可能性もある．

そこで本研究では，実際に Twitter 上で動作する非タスク指向型対話エージェントを構築し，多数のユーザとコミュニケーションを行った結果について報告する．また，エージェントの応答性能とユーザの反応についても分析し，ソーシャルメディア上でのコミュニケーションによるエージェントの自動評価の可能性を検討する．

2　対話エージェント KELDIC

本章ではユーザとのコミュニケーションを分析するため，Twitter 上に構築した対話エージェント KELDIC[5] の概要について述べる (図 1).

人間同士の対話において話し手がスムーズに話を進めていくためには，聞き手の反応や働きかけといった支援が必要であり，対話は聞き手の積極的な参加によって成立するとされている [6]．そこで，我々はユーザの発話に対し，聞き手として適切な応答を行うことで対話を活性化することを目的として本エージェントを構築した．KELDIC はフォロワーのツイートに対し，「よかったね」や「難しいね」のような短い応答を返すことで対話を進める．本エージェントは，アカウント名「@KELDIC」で 2012 年 2 月からユーザとの対話を開

*連絡先： 広島市立大学大学院情報科学研究科
〒 731-3194　広島市安佐南区大塚東 3-4-1
E-mail: inaba@hiroshima-cu.ac.jp

図 1: 対話エージェント KELDIC

始しており，本研究では開始から現在までのデータを分析する．

KELDIC の応答の流れは以下の通りである．KELDIC は 10 分に 1 回の頻度で起動し，自分のフォロワーからランダムで 50 人を抽出する．次に，抽出したフォロワーが最後に行ったツイートから，以下の 3 点をすべて満たすツイートを取得する．

- ツイートに宛先が存在しないこと (KELDIC が宛先の場合は除く)
- ツイートが投稿されてから 1 時間以内であること
- URL・画像が含まれていないこと

以上の手順により取得されたツイートに対し，KELDIC は応答を実行する．なお，実際に一回の起動で応答対象となるツイートは平均 2〜3 件であり，最大でも 10 件程度である．次節では，具体的な応答内容の決定方法について述べる．

2.1 多クラス分類に基づく応答

KELDIC は多クラス分類に基づく応答手法を採用している．すなわち，ユーザのツイートを入力，それに対する適切な応答を出力クラスとし，多クラス分類器によって応答を決定する．本研究では，多クラス分類器として多クラス SVM を用いる．そのために，あらかじめ出力となる応答クラスを決定しておく必要がある．本研究では表 1 に示した 44 種類の応答クラスを用いる．なお，応答にバリエーションを持たせるため，応答は分類器の出力 (応答クラス) をそのまま使用するのではなく，例えば「すごいね」クラスであれば,「凄いね」,「すごいですね」などの応答 (以下，これらを応答クラスに対する応答表現と呼ぶ) を複数用意しておき，当該応答クラスに対する応答表現の中からランダムで選択したものを実際の応答として使用する．

表 1: 応答クラス

すごいね	ドンマイ	マジか
かわいいよね	かっこいいよね	いいですね
ありがとう	ごめんね	さすが
そうなんだ	そうみたい	そうだね
やばいよね	よかった	よかったね
大丈夫ですか	大丈夫だよ	大変だね
本当だね	楽しそう	楽しみです
確かにね	美味しいよね	羨ましいな
面白いね	頑張ってね	頑張ろう
おめでとう	それはないね	もちろん
同感です	知らなかったよ	怖いね
了解です	お疲れさま	だめだよ
本当ですか	よろしく	嬉しいな
笑えるね	辛いね	頑張るよ
わかるよ	難しいね	

2.1.1 学習データの取得

分類器の適切な学習を行うためには，大量の学習データが存在することが望ましい．そこで，分類器を学習するための学習データとして，Twitter におけるツイート・リプライペアを用いる．すなわち，Twitter API を用いて，前述の応答クラスで Twitter を検索し，応答クラスを含むリプライと，そのリプライ先のツイートをペアとして収集する．これにより，各応答クラスがどのような発話に対して選択されるべきかというデータが取得できる．ただし，取得できるデータ数を増やすため応答表現も検索クエリとする．

手順としては，まず応答表現による検索を行い，その結果取得できたツイートから以下の条件をすべて満たすもののみを抽出する．

- 宛先のツイートが存在すること (リプライであること)
- 応答表現が文頭に存在すること
- 宛先のツイートが取得可能であること

そして，抽出したリプライの宛先となるツイートを取得する．こうして，取得したツイートを入力，応答表現の属する応答クラスを正解とする学習データが取得できる．

2.1.2 その他の機能

本稿では分析対象とはしないが，KELDIC のその他の機能として，(宛先の無い) 通常のツイートを行う機能がある．ツイートの内容は，以前我々が提案した発

図 2: 評価用 Web サイト

話候補獲得手法 [7] を用いて獲得した発話である．本手法は，入力した任意の話題語を含み，かつ 1 文で意図が理解可能な文を Twitter データから取得する手法である．KELDIC は 2 時間に 1 回，タイムライン上のユーザのツイートからランダムに話題語を選択し，発話候補獲得手法によって獲得した発話をツイートする．

その他，KELDIC をフォローしたユーザを自動でフォローする機能や，非公式 RT を用いた応答を行う機能なども有する．

2.2 応答性能

2.2.1 評価用 Web サイト

本節では，KELDIC の応答性能について述べる．

KELDIC は提案手法による応答を行う際，1%の確率で応答の最後に評価用 Web サイトへの URL を付加する．本サイトは，ユーザにより KELDIC の応答が適切であったか否かを入力してもらうことで，フィードバックを受けることを目的として構築したものである．

ユーザが KELDIC の応答に含まれる URL をクリックすると，構築した Web サイト「KELDIC の勉強部屋」へと遷移する．図 2 にそのページの画面の一例を示した．KELDIC の勉強部屋では，中央部にユーザのツイートとそれに対する KELDIC の応答が表示されており，下部に「褒める」ボタンと「叱る」ボタンが配置されている．ユーザは，KELDIC の応答がユーザのツイートに対して自然な応答であった場合は「褒める」ボタンを，不自然な応答であった場合は「叱る」ボタンをクリックすることで，KELDIC の応答を評価することができる．

表 2: 性能評価結果

	正解率 (正解数 / 総評価数)
2014 年 1〜5 月	56.6% (1019/1786)
2014 年 6〜10 月	64.4% (350/543)

2.2.2 評価結果

KELDIC は，ユーザへの応答と 2.1.1 節で述べた学習データの収集を平行して行っている．学習データの収集と再学習は 2014 年 1 月に新たに実装した機能であり，それまでは学習データは同一のものを使用していた．したがって，2014 年 1 月の応答開始以降，学習データは増加し続けている．

本機能実装前および 2014 年 1 月の開始時のデータ数 (tweet と応答クラスのペア数) は 43364 個であるが，11 月 12 日現在は 166140 個であり約 4 倍になっている．なお，KELDIC は収集した学習データを使用し，毎日深夜から早朝にかけて SVM の再学習を行っている．そこで，応答の評価結果を 1 月から 5 月と，6 月から 10 月までの 2 つの期間に区切って集計した．

構築した Web サイト上におけるユーザによる KELDIC の応答評価結果を表 3 に示す．表より，学習データが増加したため，時間の経過とともに性能が向上しており，現在は約 6 割の確率で適切な応答を返すことが可能であることが確認できた．

3 コミュニケーション分析

3.1 分析対象データ

本章では，構築した対話エージェント KELDIC とユーザのコミュニケーションの分析を行う．分析対象とするデータは，2011 年 2 月 16 日から 10 月 31 日までの KELDIC の応答とした．当該期間の間，2644 人のユーザに対し，25082 件の応答を行っており，これを分析する．ただし，ほぼすべての応答に返信を行うユーザや ID に「bot」という文字列を含むユーザは自動で応答を行う bot である可能性が高いため，人手で確認した上で分析対象から除外している．各応答には，リツイートされた数やお気に入りに登録された数などの情報も付与されており，分析にはこれらの情報も用いる．

3.2 ユーザの反応の分布

まず，KELDIC の行った応答 (ユーザのツイートに対する応答) に対するユーザの反応の分析を行った．Twitter では，他者のツイートに対する反応として，応答を

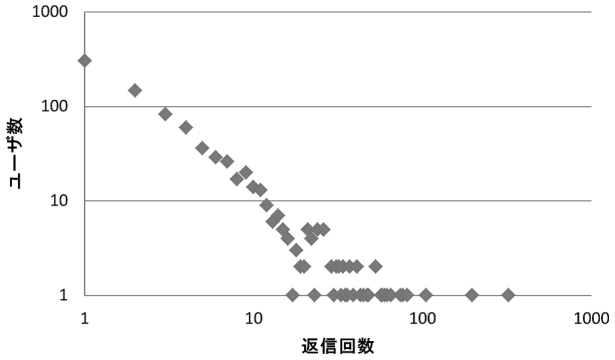

図 3: 返信回数の分布

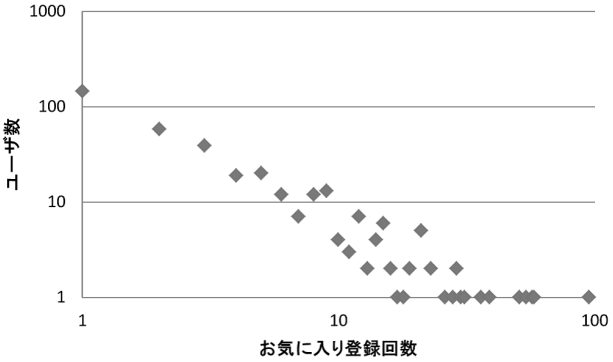

図 4: お気に入り登録回数の分布

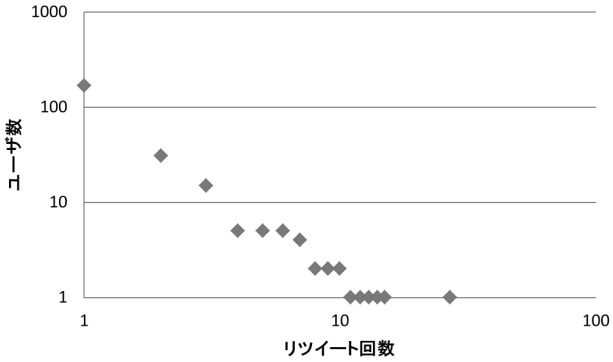

図 5: リツイート回数の分布

返す「返信 (リプライ)」のほか，ユーザが気に入ったツイートをいつでも見られるように登録する「お気に入り」と，自分のフォロワーに対して対象のツイートを知らせることができる「リツイート」がある．それぞれの反応は任意のユーザが任意のツイートに対して行うことができるが，本論文では，KELDIC が応答を行ったユーザが行った反応のみを分析対象とし，そうでないものは除外する．

まず返信に着目すると，分析対象の応答 25082 件のうち，ユーザから返信があったものは 5287 件 (返信率 21.1%) であった．ユーザ別に見ると，最も返信を多く行ったユーザの返信回数は 322 回，ユーザごとの平均返信回数は 2.0 回，その標準偏差は 9.6 であった．ただし，返信回数が 0 回のユーザは 1803 人 (68.2%) 存在し，返信回数が 1 回以上のユーザの平均返信回数は 6.3 回，標準偏差は 16.2 であった．図 3 にユーザごとの返信回数の分布を示す．横軸が返信回数，縦軸がユーザ数である．図より，返信数の分布がベキ分布となっていることがわかる．すなわち，ほとんどのユーザは全く返信を行わない，もしくは数回だけエージェントに返信を行うが，一部のユーザは 100 回を超える多くの返信を行っていることがわかる．

次にお気に入りであるが，全応答のうちユーザからお気に入りに登録されたものは 2036 件 (お気に入り登録率 8.1%) であった．最も多くお気に入りに登録を行ったユーザの登録数は 95 件，ユーザごとの平均登録数は 0.8 件，標準偏差は 3.9 であった．また，リツイートに関しては，全応答のうちユーザからリツイートされたものは 524 件 (リツイート率 1.0%) であった．最も多くリツイートしたユーザのリツイート回数は 27 件，ユーザごとの平均リツイート回数は 0.2 件，標準偏差は 1.1 であった．ここから，リツイートはお気に入りに対し，より少数の応答に対して行われていることがわかる．これは，お気に入りとは違い，リツイートは他のユーザのタイムラインにも影響をあたえるため，心理的敷居が高いことが影響していると考えられる．なお，最も返信を多く行ったユーザと最も多くお気に入りに登録を行ったユーザ，および最も多くリツイートしたユーザはそれぞれ別のユーザであった．

図 4 にユーザごとのお気に入り登録回数の分布を，図 5 にリツイート回数の分布を示した．これらの図より，図 3 の返信数と同じく，お気に入り登録回数とリツイート回数もベキ分布となっていることが分かる．

ここで，ユーザの反応の傾向を分析するため，反応回数について相関関係の分析を行った．その結果，お気に入り登録とリツイートのどちらかを少なくとも 1 回以上行ったユーザにおける，お気に入り登録数とリツイート回数の相関係数は 0.26 であり，弱い相関関係にあることが確認された．一方，返信数とお気に入り数・リツイート数のそれぞれの相関係数は両者とも 0.02 であり，無相関であることが確認された．ここから，エージェントに対して積極的に反応するユーザは，頻繁に返信することでやりとりを楽しむユーザと，好みの応答をお気に入り登録・リツイートをするユーザの 2 種類に分かれていることがわかった．

表 3: 応答性能とユーザの反応分析結果

	分析対象応答数	正解応答数	正解率 (%)	反応なしとの有意差
反応なし	200	131	65.5	—
返信	200	143	71.5	有意差なし
お気に入り	200	141	70.5	有意差なし
リツイート	144	100	69.4	有意差なし
返信 + お気に入り	200	164	82.0	有意水準 1%で有意差あり
返信 + リツイート	43	37	86.0	有意水準 1%で有意差あり
お気に入り + リツイート	133	92	69.2	有意差なし
返信 + お気に入り + リツイート	27	20	74.1	有意差なし

3.3 応答性能とユーザの反応

3.3.1 分析方法

対話エージェントがどの程度適切な応答が可能かという応答性能は，ユーザの反応にも影響を与えると考えられる．そこで，ユーザにより返信，お気に入り登録，リツイートされた応答，また，それらの 2 つ以上の組み合わせた反応があった応答，および全く反応が無かった応答をそれぞれ人手で確認することで，応答性能とユーザの反応の関係を分析する．反応については，各応答ごとの応答対象ユーザによるもののみを対象とし，それ以外のユーザによる反応は考慮しない．分析を行う応答は，各反応ごとに最大 200 個とする．ただし，ユーザごとの反応の偏りを排し，分析結果の一般性を確保するため，それぞれ最大 200 個の発話のうち同一ユーザが反応を行ったものは 3 個以内になるようにした．

3.3.2 分析結果

分析結果を表 3 に示した．表の正解応答数は，ユーザのツイートに対して適切な応答であった応答の数，正解率は正解応答数の分析対象の発話全体に占める割合である．また，反応なしとの有意差には，反応なしと各反応の間で比率の差の検定を行った結果について示した．表より，反応なしと比較すると，ユーザから何らかの反応があった応答の方が正解率は高いことがわかる．各反応が単体で行われた場合に着目すると，返信，お気に入り，リツイートのそれぞれの正解率にほとんど差は見られなかった．これは，返信には KELDIC の応答の意味がわからなかったことを意味する「は？」や「どういう意味ですか」のような返信も多数含まれていることや，応答の意味は正しくないが，変な応答が来たということでお気に入りやリツイートするユーザが存在していることが，正解率に差が見られなかった理由であると思われる．

2 つ以上の反応の組み合わせでは,「返信 + お気に入り」と「返信 + リツイート」の場合で,「反応なし」の場合と有意水準 1%で有意差が確認された．一方,「お気に入り + リツイート」と「返信 + お気に入り + リツイート」では有意差が確認できなかった．まず,「お気に入り + リツイート」の場合であるが，これは Twitter において，お気に入りとリツイートが同じインタフェースで実行可能であることが影響しているものと思われる．Twitter では，各ツイートの下にお気に入り登録ボタンとリツイート登録ボタンが並んでいるという UI を採用しており，どちらもボタンをクリックすることで実行可能である．また，お気に入りとリツイートの両方を行った場合とお気に入り・リツイートをそれぞれ単体で反応した場合の正解率はほぼ同じである．前節におけるお気に入り登録数とリツイート回数が弱い相関を示した結果を踏まえると，お気に入りとリツイートの区別を意識せず，両方とも登録するユーザが一定数存在しているものと考えられる．

「返信 + お気に入り + リツイート」の有意差が確認できなかったことについては，このようなケースが非常に少ないことが主な原因であると考えられることから，より多くのデータを収集した後，再度分析が必要である．

以上の分析結果から，返信に加え，お気に入り登録かリツイートが行われた応答は正解である割合が他と比較して大きいことが明らかになった．これは前節の分析から，ユーザは返信を好むユーザとお気に入り登録・リツイートを好むユーザに分かれており，それらを同時に行うということは，そのエージェントの応答がユーザにとって特別であった場合に限られているためと考えられる．一方で，それぞれの反応が単体で行われた場合には，反応がない場合よりは正解率が高くなったが，統計的な差までは確認できなかった．

表 4: ユーザの反応による応答評価

	反応ありの応答数	反応なしの応答数	反応ありの割合
2014 年 1〜5 月	75	3479	2.1%
2014 年 6〜10 月	294	5113	5.4%

4 応答性能自動評価の可能性

本章では，前章における分析結果を踏まえ，ユーザの反応を用いた対話エージェントの応答性能評価が可能であるかを検討する．性能比較のため，表 2 に示した応答時期の異なる KELDIC 同士の比較を行う．これは，応答内容が大きく異なるエージェント同士の場合，ユーザの返信率なども大きく差が生じると考えられるため，別個のエージェント間の比較は困難であると予想されるためである．

ユーザの反応を用いた評価は，前節の分析結果を踏まえ「返信 + お気に入り」もしくは「返信 + リツイート」のどちらかが行われた応答 (反応ありの応答) と，反応なしの応答を対象とし，反応ありの応答数の割合の比較を行う．

評価結果を表 4 に示す．表より，1 月〜5 月よりも 6〜10 月の方が反応ありの割合が高いことが確認でき，6〜10 月の方が性能が高いことを示唆している．またこの結果は表 2 とも一致している．さらに，1 月〜5 月と 6〜10 月の反応ありの割合で比率の差の検定を実施した結果，有意水準 1%で有意差が確認された．以上の結果から，ユーザの反応を用いることで，応答の自動評価が行える可能性が示唆された．

しかし,「返信 + お気に入り」もしくは「返信 + リツイート」が行われる応答は非常に少ないため．この評価方法では信頼できる評価を行うためにはかなりの時間を要することから，効率が悪いことも明らかとなった．したがって，応答の頻度を増やしたり，対話するユーザ，すなわちフォロワーを増やすなどの対策が必要である．また，本稿では応答の意味的な正しさという評価基準を用いたが，ユーザから反応がある応答は内容の面白さなど，自然さ以外の要素も大きいと考えられるため，ユーザの反応と応答内容の関係に関してより詳細な分析も必要であると思われる．

5 まとめ

本研究では，Twitter 上で動作する非タスク指向型対話エージェント KELDIC を構築し，本エージェントとユーザとのコミュニケーションの分析を行った．Twitter におけるユーザのコミュニケーション方法には，返信，お気に入り登録，リツイートの 3 種類があり，それぞれの反応の統計的特徴を明らかにした．また，ユーザの反応を用いることで，エージェントの応答性能の評価できる可能性が示唆された．

今後は，より複雑な応答が可能なエージェントを Twitter 上で動作させ，より詳細な分析を進めることや，ユーザの反応によるエージェントの自動学習の可能性についても検討していきたい．

参考文献

[1] Hiroaki Sugiyama, Toyomi Meguro, Ryuichiro Higashinaka, and Yasuhiro Minami. Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In *Proc. SIGDIAL*, pp. 334–338, 2013.

[2] Rafael E Banchs and Haizhou Li. Iris: a chat-oriented dialogue system based on the vector space model. In *Proceedings of the ACL 2012 System Demonstrations*, pp. 37–42. Association for Computational Linguistics, 2012.

[3] Alan Ritter, Colin Cherry, and Bill Dolan. Unsupervised modeling of twitter conversations. *In Proc. NAACL-HLT*, pp. 172–180, 2010.

[4] NIO Lasguido, Sakriani SAKTI, Graham NEUBIG, TODA Tomoki, and Satoshi NAKAMURA. Utilizing human-to-human conversation examples for a multi domain chat-oriented dialog system. *IEICE TRANSACTIONS on Information and Systems*, Vol. 97, No. 6, pp. 1497–1505, 2014.

[5] 稲葉通将, 高橋健一. Twitter から学習する対話エージェントの設計. 合同エージェントワークショップ＆シンポジウム 2014, 2014.

[6] 堀口純子. 日本語教育と会話分析. くろしお出版, 1997.

[7] 稲葉通将, 神園彩香, 高橋健一. Twitter を用いた非タスク指向型対話システムのための発話候補文獲得. 人工知能学会論文誌, Vol. 29, No. 1, pp. 21–31, 2014.